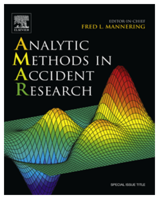
Analytic Methods in Accident Research 9 (2016) 16–26

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

journal homepage: www.elsevier.com/locate/amar



Evaluating crash type covariances and roadway geometric marginal effects using the multivariate Poisson gamma mixture model

Ghasak I.M.A. Mothafer [a], Toshiyuki Yamamoto [b], Venkataraman N. Shankar [c,*]

[a] Department of Civil Engineering, Nagoya University, Nagoya 464-8603, Japan
[b] EcoTopia Science Institute, Nagoya University, Nagoya 464-8603, Japan
[c] Department of Civil and Environmental Engineering, Penn State University, PA 16802, USA

ARTICLE INFO

Article history:
Received 30 September 2015
Received in revised form
6 November 2015
Accepted 7 November 2015
Available online 17 December 2015

Keywords:
Multivariate count data
Poisson gamma mixture
Crash types
Covariances

ABSTRACT

This paper investigates the correlations and covariances among the rear end, sideswipe, fixed object and other crash types on freeway sections using three-year crash data for 274 multilane freeway segments in the State of Washington, U.S.A. A multivariate Poisson gamma mixture count model (MVPGM) is developed assuming positive correlation among crash types. The model parameters are estimated using a maximum likelihood approach. Based on the empirical results, the proposed model shows significant unobserved correlations among different types of crash frequencies. In addition to evaluating crash type correlations and covariances by crash type, the model also allows for evaluation of roadway geometric marginal effects and how they compare with crash type-specific effects. The results show that the MVPGM covariances of crash types are in better agreement with observed covariances than those from univariate crash type models. These findings are in spite of our observation that the individual crash type models provide for statistically better fits due to their unconstrained dispersion parameters, which are constrained to be the same in the multivariate model we have proposed here. This outcome underscores the need to explore the behavior of dispersion in multivariate crash type contexts.

1. Introduction

An enormous body of literature has been devoted to modeling crash and safety considerations. Lord and Mannering (2010) show an extensive review and assessment of methodological alternatives related to crash-frequency data. Crash frequency and severity are two key indices that measure safety risk for a roadway segment. The classical approach uses a conventional frequency model to predict crash counts as univariate dependent variables. For example Lord et al. (2005) have compared Poisson, Poisson-Gamma and zero-inflated regression models of motor vehicle crashes. The use of crash type as an independent variable is typically restricted to the analysis of crash severity (Chiou and Fu, 2013; Shaheed et al., 2013;

* Corresponding author.
E-mail addresses: mothafer.g.i.m.a@c.mbox.nagoya-u.ac.jp (G.I.M.A. Mothafer), yamamoto@civil.nagoya-u.ac.jp (T. Yamamoto), shankarv@engr.psu.edu (V.N. Shankar).

http://dx.doi.org/10.1016/j.amar.2015.11.001
2213-6657/

Gkritza et al., 2010; Yan et al., 2011; Yang et al., 2011; Shankar et al., 1996); while other studies have comprehensively examined a particular type of crash at a specific facility (Das and Abdel-Aty, 2011; Dissanayake and Lu, 2002).

The need for developing and estimating crash type occurrence propensities jointly has been recognized in the literature (Ma et al., 2008; Park and Lord, 2007). Several researchers have investigated freeway crash statistics and attempted to develop models of crash frequencies for freeways. Ye et al. (2013) model crash frequency by severity on freeways using the simultaneous equations Poisson log normal model with error components that are normally distributed. Anastasopoulos et al. (2012) have utilized the multivariate tobit model by considering crash rates instead of crash frequencies. Chiou and Fu (2013) have modeled crash frequency by severity using the multinomial-generalized Poisson model with error components, while Dong et al. (2014) have used multivariate random-parameters zero-inflated negative binomial regression models to estimate crash frequencies of different types at intersections.

In general, there are five multivariate count model types to estimate the correlation among frequencies by crash type or severity: the multivariate Poisson model; the multivariate negative binomial model; the multivariate Poisson-gamma mixture model; the multivariate Poisson-log-normal model and the latent Poisson-normal model. Multivariate Poisson-log-normal models have been used extensively in the literature for both crash types and severity. Ye et al. (2013) and Chiou and Fu (2013) have used the maximum simulated likelihood method to estimate the parameters of this model, while Park and Lord (2007), El-Basyouny and Sayed (2009), Ma et al. (2008) and Ma and Kockelman (2006) have used the Bayesian approach. Nevertheless, there is some doubt whether these models could be applied to high dimensional multivariate data effectively (Winkelmann, 2008). The other model namely the multivariate latent Poisson-normal incorporates a non-linear parameterization of the thresholds as a function of exogenous variables (Castro et al., 2013; Castro et al., 2012; Yasmin et al., 2014; Bhat et al., 2014). Complexity and non-closed-form of the joint probability density are considered as limitations to the estimability of these models.

The Poisson gamma mixture model was first introduced by Hausman et al. (1984) with further explanation of its use by Dey and Chung (1992). Miles (2001) provides an application of this model to individual consumer data on the number of purchases of bread and cookies in a one-week period, using maximum likelihood estimation (MLE) of the Poisson gamma mixture probability. In the Hausman-et al/Miles models, correlation is generated by an individual specific multiplicative error term. Kockelman (2001) conducted a time and budget constrained activity demand analysis utilizing the same model as Miles (2001). This model offers a closed form and is easy to estimate by using the maximum likelihood method.

This paper proceeds on similar lines by developing a multivariate Poisson gamma mixture (MVPGM) model to jointly model crash frequencies by type considering the effects of various roadway, geometric and traffic volume factors on crash frequencies. The focus of this paper however is on the nature of covariances among crash types. Understanding covariance among crash types is an important step in the development of safety interventions – it potentially improves the ability to target multiple crash types through fewer interventions, thereby enhancing the cost effectiveness of safety improvements. To this end, as a methodological first step, the proposed model considers the covariance matrix through error components specified under an integrated model framework among crash types. Model estimation is achieved through the use of the maximum likelihood estimation method.

The rest of the paper is structured as follows. The next section presents a description of key variables pertaining to the various crash types, including information on crash type distributions, roadway geometrics and traffic volume. Section 3 discusses model formulation and implications for correlation across crash types. Section 4 illustrates an application of the proposed model for analyzing crash type counts on interstates, and includes a discussion of covariances and marginal effects with respect to the joint model. The fifth and final section offers concluding thoughts and directions for further research.

2. Data

The crash dataset is obtained for interstate 5 (approximately 276 miles in length) in the State of Washington, USA. Three years of crash data were collected for the period 2005–2007. Data contained three different categories: (1) crash frequencies aggregated from individual crash data; (2) geometric characteristics; and (3) traffic volume information. Individual crash data included information on crash type and the most severe outcome of the crash. An observation in the dataset either represented an interchange segment or a non-interchange segment. Interchange segments were defined as segments bounded by the farthest ramp terminal on either side of an interchange overpass. Noninterchange segments were segments located between two interchange segments. Average segment length was roughly 0.87 miles with a standard deviation of about 0.60 miles. For each segment, crashes were classified by year, resulting in three years of data per segment.

In total, 13,357 crashes, including 7,512 rear-end, 2,377 sideswipe, 2,004 fixed object, 242 overturn, 685 same-direction, 12 head-on and 525 miscellaneous crash types were analyzed. Crash frequencies by type and year are shown in Table 1. On state highways, the major collision types usually are: same direction collisions resulting in rear end or sideswipe crashes, as well as fixed objects and entering at angle. The proportion of entering at angle collisions on interstates, especially on the mainline is virtually negligible, with the result that three major collision types dominate the frequency distribution. This is confirmed by Table 1, which shows that rear-end, sideswipe, and fixed objects crash types comprise a large proportion of the frequency distribution of crash types, with the remaining category classified as "all-other," and inclusive of same